


# Nordic adolescents responding to demanding survey scales in boring contexts: Examining straightlining

Leif Edvard Aarø<sup>1,2</sup>  | Anne-Siri Fismen<sup>1,3</sup> | Bente Wold<sup>2</sup> |  
 Jens Christoffer Skogen<sup>1,3,4</sup> | Torbjørn Torsheim<sup>5</sup> | Ársæll Már Arnarsson<sup>6</sup> |  
 Nelli Lyyra<sup>7</sup> | Petra Löfstedt<sup>8</sup> | Charli Eriksson<sup>9</sup>

<sup>1</sup>Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway

<sup>2</sup>Department of Health Promotion and Development, University of Bergen, Bergen, Norway

<sup>3</sup>Centre for Evaluation of Public Health Measures, Norwegian Institute of Public Health, Oslo, Norway

<sup>4</sup>Alcohol and Drug Research Western Norway, Stavanger University Hospital, Stavanger, Norway

<sup>5</sup>Department of Psychosocial Science, University of Bergen, Bergen, Norway

<sup>6</sup>Faculty of Health Promotion, Sport and Leisure Studies, University of Iceland, Reykjavík, Iceland

<sup>7</sup>Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>8</sup>Department of Public Health and Community Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

<sup>9</sup>Department of Learning, Informatics, Management, and Ethics, Karolinska Institute, Stockholm, Sweden

## Correspondence

Leif Edvard Aarø, Department of Health Promotion, Norwegian Institute of Public Health, Zander Kaes gt. 7, 5015 Bergen, Norway.  
 Email: leea@thi.no

## Funding information

Folkhälsomyndigheten

## Abstract

**Introduction:** Straightlining, or identical responses across all items within a multi-item scale, is often taken as an indication that responses to all items in a questionnaire are of poor quality. The purpose of this study was to examine straightlining on two scales: The Sense of Unity Scale (SUS) and the short version of the Warwick-Edinburgh Mental Well-being Scale (SWEMWBS).

**Methods:** Data stem from the 2017–2018 data collections in four Nordic countries of the Health Behaviour in School-children study (HBSC) (15-year-old students only; 50.9% girls;  $n = 5928$ ). Data were weighted to adjust for oversampling of Swedish-speaking Finnish students and to equalize sample size across countries. The main analyses were done with general linear modeling with adjustments for cluster effects (school classes).

**Results:** The proportion with straightlining on SUS was 22.8%, varying from 5.8% among Swedish girls to 46.4% among Finnish boys. The proportion with straightlining on SWEMWBS was 18.4%, varying from 5.2% among Norwegian girls to 46.0% among Finnish boys. Straightlining on one of the scales correlated with straightlining on the other one. Straightlining tended to inflate Cronbach's  $\alpha$  values and reduce number of factors in factor analyses. Associations between the two scales and external variables tended to be lower among straightlining students. Associations between external variables (other than SUS/SWEMWBS) are on average slightly weaker among straightliners. Straightlining students obtained more favorable scores on several resource-related variables.

**Conclusion:** Although some problems have been identified, straightlining does not serve well as a general indicator of poor data quality.

## KEYWORDS

adolescents, Nordic countries, questionnaires, scales, straightlining, surveys

## 1 | INTRODUCTION

Research on the development of adolescent mental health and quality of life is challenging and studies using high-quality data are crucial. This has been the emphasis in a collaboration among researchers from the Nordic countries, utilizing data from the 2017/2018 data collection of the Health Behaviour in School-aged Children study (HBSC) (Eriksson et al., 2019).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Adolescence* published by Wiley Periodicals LLC on behalf of Foundation for Professionals in Services to Adolescents.

This paper focuses on a kind of response set called “straightlining” in relation to two measures of positive mental health: the Sense of Unity Scale (SUS) and the short version of Warwick-Edinburgh Mental Well-being Scale (SWEMWBS).

“Response bias” refers to factors or conditions that take place during the process of responding to survey questions and which affect the responses and thereby lead to a nonrandom deviation of the answers from their true value (Villar, 2008, p. 751). A subcategory of response biases is “response styles,” defined as “a respondent's tendency to respond to survey questions in certain ways regardless of the content” (Van Vaerenbergh & Thomas, 2013). The definition of “response set” is strikingly similar. There are several kinds of response sets described in the scientific literature, for instance, the acquiescence bias, the bias of extreme responding, bias to the middle, random responses, and straightlining.

The occurrence of response bias is related to the length of the questionnaire and types of questions and response categories. If respondents find the questions irrelevant or boring, this might produce less thoughtful answers. This is particularly so if the questions come towards the end of a long questionnaire (Herzog & Bachman, 1981). Adolescence is a period of life characterized by a high level of impatience and impulsivity (Steinberg et al., 2009). Fatigue and boredom when filling out questionnaires may therefore be particularly challenging in surveys among adolescents.

A simple and broadly accepted remedy for at least some types of response sets was suggested by Moser and Kalton. “Variation between positive and negative items forces the respondent to consider each item carefully, rather than to respond automatically to them all in the same way” (Moser & Kalton, 1971, p. 362). Changing the order of response categories across a set of items introduces new problems. One challenge is that it becomes more demanding and time consuming to respond to the questions. In many, perhaps most multi-item scales, the order of response categories is therefore the same for all items.

Instead of investing the necessary efforts into providing thorough answers, study participants may simply provide answers that appear to be adequate. Krosnick (1991) coined the term “satisficing” to characterize this kind of responses. Krosnick pointed at six different strategies of “satisficing”: (i) selecting the first response alternative that seems reasonable, (ii) agreeing with assertions, (iii) endorsing the status quo (not particularly relevant in our context), (iv) non-differentiation in using rating scales, (v) saying “Don't know,” and (vi) mental coin-flipping.

Krosnick hypothesized that the risk of satisficing increases with the difficulty of the scale and decreases with study participants' ability and motivation. To the extent that the scale is cognitively demanding, and study participants are unmotivated and lack cognitive abilities, satisficing responses will be more common.

“Non-differentiation” or “straightlining” in using rating scales is defined as selecting a response category on the first item, and then using the same response category for all remaining items (Krosnick, 1991; Loosveldt & Beullens, 2017).

Straightlining has been found to be more common among respondents with low levels of education and among respondents who have low levels of verbal abilities, and straightlining is more common towards the end of questionnaires (Herzog & Bachman, 1981; Krosnick, 1991).

Some researchers, for instance Yan (2008), regard straightlining to be a source of measurement error which contributes to decreasing quality of data. Reuning and Plutzer (2020) are, however, critical of the practice of regarding the occurrence of straightlining as an indicator of low study quality. They are also critical of the practice of removing respondents identified as straightliners from data based on the assumption that their answers are meaningless. They distinguish between valid versus invalid straightlining: “Valid straightlining occurs when two conditions are satisfied: A respondent is motivated to carefully read/listen to questions and answer them truthfully, and provides identical (non-differentiated) responses to a sequence of questions” (Reuning & Plutzer, 2020, p. 440). Invalid straightlining occurs when items are not examined carefully, and responses are provided without sufficient attention and cognitive processing.

The likelihood that a study participant will provide identical responses to all items decreases with increasing numbers of items and response alternatives. In the 2017/2018 data collection of the WHO Collaborative Study on Health Behaviour in School-aged Children (the HBSC Study), there were two scales that we will use for a closer inspection of straightlining—a seven items version of the Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) (Ringdal et al., 2017; Stewart-Brown et al., 2009) and the Sense of Unity Scale (SUS) which comprises eight items (Eriksson et al., 2019; Samdal et al., 2016). All items in both scales have five response categories. The response categories and their order are identical within scales. Both scales are relatively abstract and cognitively demanding. Since identical scales were used across four countries, the total number of observations was sufficient for the analyses carried out in this study. The data at hand are therefore rather ideal in a study of straightlining.

As mentioned above, with reference to Reuning and Plutzer (2020), identical responses across all items are not necessarily an indication of bias. A young person may examine each item of the SWEMWBS scale items carefully and come to the exact same conclusion for each item, for instance that the answer is always “Some of the time,” which is the third response category out of five possible. If the proportion of students with this kind of uniform answers across items is very high, or varies strongly across groups of students, this may indicate bias at the group level. Associations with third variables which deviate strongly from similar associations among those who have not practiced any straightlining at all may be another indication of response bias. If straightlining means that responses are arbitrary (characterized by some degree of randomness), lower correlations between scores on the scale where straightlining has taken place and other variables can be expected.

The purposes of the present paper are to (i) examine the occurrence of straightlining among students in four Nordic countries (Finland, Iceland, Norway, and Sweden) and (ii) differences in the prevalence of straightlining across genders and countries. We will also (iii) examine effects of straightlining on the dimensionality and internal consistency of scales, and (iv) examine differences in associations with external variables among straightliners and non-straightliners. Finally, we want to (v) describe differences between straightliners and non-straightliners on several resource-related variables and (vi) compare intercorrelations and associations among external variables among straightliners and non-straightliners.

For researchers using survey methodology straightlining is an important issue. If straightlining is a strong indicator of poor quality of survey responses, steps should be taken to reduce the risk of straightlining when new data are collected. And when data are analyzed, examining effects of straightlining on data quality may turn out to be imperative.

## 2 | MATERIALS AND METHODS

### 2.1 | The HBSC study

This study is based on data from the HBSC, a World Health Organization (WHO) collaborative cross-national survey with an overall aim to generate increased understanding of health, health behavior, and their contexts in the lives of young people aged 11, 13, and 15 years. The present study includes samples of 15 years old Finnish ( $n = 1146$ ), Icelandic ( $n = 2195$ ), Norwegian ( $n = 874$ ), and Swedish ( $n = 1606$ ) students participating in the HBSC study for the survey year 2017/2018.

### 2.2 | Samples and data collection

In Finland, Norway, and Sweden, samples were drawn randomly by cluster sampling within strata defined by, for instance, grade, geographical area, and language. School classes were used as sampling units, and stratification was proportional. The recommended national sample size per grade was 1500, and the mean age should be 11.5, 13.5, and 15.5 years, where 90% of the sample should fall between  $\pm 6$  months of the mean age. Iceland invited all schools in the country and therefore had a larger sample compared to the other countries. As mentioned above, the present study is based on data from the oldest age group (15 years) only.

In the HBSC study the English version of the questionnaires is the primary one. Procedures for translation and adaption to other languages have been established. This includes back-translations and piloting. Questionnaires were made available in Finnish and Swedish in Finland, Swedish in Sweden, in the two versions of Norwegian—Norwegian “bokmål” and Norwegian “nynorsk”—in Norway, and Icelandic in Iceland. The questionnaire was not administered in any of the immigrant group languages or in the Sami language. The Swedish language version of the questionnaire in Finland was not identical to the Swedish language questionnaire used in Sweden, but simply a Swedish language version of the Finnish language questionnaire. The two Norwegian languages are too similar to represent any challenge with regard to translations.

The students answered the questionnaire at school, after receiving instructions from their teacher. Oral and written information on the confidentiality of their responses were provided and participation was confidential and voluntary.

In Finland, Iceland, and Norway, the questionnaire was web-based. In Sweden they used a printed questionnaire for data collection. Data collections took place in class during school hours. Data collections could take place any day of the week, from Monday to Friday, and during any school hour, but the day and hour had to be the same for all students within one school class.

A standardized international research protocol was followed to ensure consistency in survey instruments, data collection, and processing procedures. Response rates for schools/students varied across the countries: Finland 47/60%, Norway 15/71%, and Sweden 47/89%. In Iceland the response rate for the whole student population was 49%, but as the largest municipality in the country, Reykjavik, did not allow tracking individual schools, it was impossible to estimate a separate response rate for schools (Inchley et al., 2020). Schools or classes that declined to participate, as well as students absent on the day the survey was carried out, were the two main sources of nonresponse and were not followed up. The items used in the present study were part of optional HBSC packages, included by Nordic countries and part of a Nordic HBSC collaboration on mental health (Eriksson et al., 2019).

The HBSC Data Management Centre, located at the University of Bergen, Norway, checked the quality of the data collected, performed appropriate cleaning of the data, and merged national data sets into an international data file. Detailed information about the study is available at <http://www.hbsc.org/>. The methodology for data collection is described in the HBSC protocol (Inchley et al., 2018), which prescribes consistency in sampling plans, survey instruments, and data collection.

## 2.3 | Measures

The *Sense of Unity Scale* was developed within the HBSC study and refers to the positive feeling of being part of a larger social structure (Wahlström et al., 2021). The measure was piloted in Norway in 2016 and included as part of an optional package in the international survey in 2017/2018. The pilot study reports a high internal consistency (Cronbach's  $\alpha = .89$ ; Samdal et al., 2016) and a qualitative study in the form of a classroom discussion (28 students aged 13–14 years) about all the items indicated high validity and good understanding of the items.

The Sense of Unity scale comprises eight items: “I feel...”; “a strong sense of togetherness”; “responsibility for others”; “that it is good to be part of a community”; “that I contribute without expecting anything in return”; “that I mean a lot to others”; “that others mean a lot to me”; “that I trust most people”; and “that we rely upon each other.” The response alternatives were “Not at all true for me,” “Not very true for me,” “Neither true nor untrue for me,” “Somewhat true for me,” and “Really true for me.” The wording of the response categories varied slightly across the four countries. Categories were coded 1 to 5, and these scores were summed to an index ranging from 8 to 40, higher values indicating higher sense of unity. In the analyses presented in this publication, a mean score was used instead, the sum score was divided by the number of items (range 1–5).

In the context of this study, it is important to notice that in three countries (Finland, Iceland, and Norway), the SUS had a rather wordy introduction: “Everyone belongs to different social groups and communities such as school classes, neighborhoods, sports teams, towns, countries, and wider society. Rate how true each statement regarding how you feel about belonging to such groups and communities is for you ranging from ‘Not at all true for me’ to ‘Really true for me’. Mark the box under the rating that best applies to you. I feel ...”. In the Swedish questionnaire most of this introduction was removed. Only the words “I feel ...” (“Jag känner ...”) remained.

In the HBSC data collections in the countries included in this study, the seven-item version of the Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) was used: “I’ve been feeling optimistic about the future,” “I’ve been feeling useful,” “I’ve been feeling relaxed,” “I’ve been dealing with problems well,” “I’ve been thinking clearly,” “I’ve been feeling close to other people,” and “I’ve been able to make up my own mind about things.” The response alternatives were: “None of the time,” “Rarely,” “Some of the time,” “Often,” and “All of the time” (Tennant et al., 2007). Categories were coded 1–5, and a mean score with higher values indicating higher levels of mental well-being was constructed.

Straightlining was in this study defined as giving the same response across all items within one scale. This means that straightlining could only be detected among those who had valid answers on all scale items.

A number of questions and scales served as external variables which were analyzed against SUS and SWEMWBS. Analyses of associations among these external variables were also carried out. Most of these variables belong to the domains of mental health, quality of life, or interpersonal relationships. For each multi-item scale described below, mean scores were constructed.

*Life Satisfaction* was a single item measure: “Here is a picture of a ladder. The top of the ladder ‘10’ is the best possible life for you and the bottom ‘0’ is the worst possible life for you. In general, where on the ladder do you feel you stand at the moment? Tick the box next to the number that best describes where you stand.” The students were instructed to select one of the boxes which were drawn between the steps of a ladder with zero at the bottom and ten at the top (Cantril, 1965).

Previous research has shown good test–retest reliability and validity of this scale among adolescents (Jovanović, 2016; Levin & Currie, 2014).

*Self-esteem* was measured with a scale based on Rosenberg’s general self-esteem model. It has been used in the Danish Methodological Study (Nielsen et al., 2016). Three items are used to measure students’ positive conceptions of self and their conceptions of what others think about them: “I like myself,” “I am good enough as I am,” and “Others my age like me.” Five response categories were used: “Strongly agree,” “Agree,” “Neither agree nor disagree,” “Disagree,” and “Strongly disagree.” Categories were coded 1 to 5 with higher values reflecting high self-esteem.

*General Self-efficacy* was a two-items measure: “How often do you find a solution to a problem if you try hard enough?” and “How often do you manage to do the things that you decide to do?”. Response categories were: “Never,” “Rarely,” “Sometimes,” “Most of the time,” “Always,” coded 1 to 5. (Schwarzer, 1992). This short version has been used in a series of studies (Meilstrup et al., 2020; Nielsen et al., 2016).

*Loneliness* was assessed using a single item on global loneliness. In all countries other than Sweden, loneliness was assessed by the question ‘Do you feel lonely?’ with four response categories: 1 “Yes, very often,” 2 “Yes, quite often,” 3 “Sometimes,” and 4 “No.” In Sweden, adolescents were asked ‘How often do you feel lonely?’ with five response categories: 1 “Always,” 2 “Most of the time,” 3 “Sometimes,” 4 “Rarely,” and 5 “Never.” For Swedish data, the response categories 4 and 5 were combined to better match the response categories used in other Nordic countries. The direct, single-item measure of loneliness is highly correlated with the multi-item UCLA Loneliness Scale, which is an indirect measure of loneliness (Lyyra et al., 2021). The response categories were recoded in order to let high values indicate high levels of loneliness.

*Health complaints* were measured with the HBSC Symptoms Checklist. “In the last 6 months: how often have you had the following?” The list of complaints included Headache, Stomachache, Backache, Feeling low, Irritability or bad temper,



Feeling nervous, Difficulties in getting to sleep, and Feeling dizzy. The response alternatives were: “Rarely or never,” “About every month,” “About every week,” “More than once a week,” “About every day” (Hetland et al., 2002). The categories were coded 1 to 5 with higher numbers indicating higher levels of complaints.

*Self-rated global health* was measured with one item: “Would you say your health is ...? Four response alternatives were provided: “Poor,” “Fair,” “Good,” “Excellent.” The categories were coded 1 to 4 with higher numbers indicating better health.

*Liking school* was measured with one item: “How do you feel about school at present?” Response categories were: “I like it a lot,” “I like it a bit,” “I don't like it very much,” and “I don't like it at all.” The categories were coded 1 to 4 with higher numbers indicating stronger liking of school.

*Relationship with teachers* was measured with three items: (i) I feel that my teachers accept me as I am, (ii) I feel that my teachers care about me as a person, and (iii) I feel a lot of trust in my teachers. Response categories were: “Strongly agree,” “Agree,” “Neither agree nor disagree,” “Disagree,” and “Strongly disagree.” The categories were coded 1 to 5, with higher numbers indicating stronger agreement. The cross-country reliability and validity of the teacher support scale has been demonstrated previously (Torsheim et al., 2000, 2012).

Family support and Peer support were measured using the family and friend subscales of the Multidimensional Scale of Perceived Social Support (MSPSS) (Zimet et al., 1988). The scale is reliable and valid, as has been shown in studies among various adolescent samples from different countries (Canty-Mitchell & Zimet, 2000; Edwards, 2010).

*Peer support* was measured with four items: (i) “My friends really try to help me,” (ii) “I can count on my friends when things go wrong,” (iii) “I have friends with whom I can share my joys and sorrows,” and (iv) “I can talk about my problems with my friends.” Responses were indicated on a scale from 1 to 7 with labels on the two endpoints. “Very strongly disagree” was indicated by the number 1 and “Very strongly agree” by the number 7.

*Family support*: “We are interested in how you feel about the following statements. Please show how much you agree or disagree with each one. Please tick one box for each line.” The following four items were listed: “My family really tries to help me,” “I get the emotional help and support I need from my family,” “I can talk about my problems with my family,” and “My family is willing to help me make decisions.” Study participants could indicate on a seven-point scale how much they agreed or disagreed with “Very strongly disagree” at one end (coded 1) and “Very strongly agree” at the other end (coded 7) (Zimet et al., 1988).

*Family affluence Scale III* comprised six items: “Does your family own a car, van or truck?” (“No,” “Yes, one,” “Yes, two or more”). “Do you have your own bedroom for yourself?” (“No,” “yes”). “How many computers do your family own (including laptops and tablets, *not* including game consoles and smartphones)?” (“None,” “One,” “Two,” “More than two”). “How many bathrooms (room with a bath/shower or both) are in your home?” (“None,” “One,” “Two,” “More than two”). “Does your family have a dishwasher at home?” (“No,” “Yes”). “How many times did you and your family travel out of [insert country here] for a holiday/vacation last year?” (“Not at all,” “Once,” “Twice,” “More than twice”) (Currie et al., 1997; Hartley et al., 2016). A riddit transformation (conversion to cumulative probabilities) by age and sex was carried out on the sum score of the Family Affluence Scale III items.

## 2.4 | Statistical analyses

Data were analyzed with SPSS Version 27. Since school classes were sampling units, adjustments for cluster effects were necessary. Analyses of two-way frequency tables were carried out in SPSS Complex. Analyses of dimensionality were done with factor analysis with principal axis factoring, pairwise deletion of cases and direct oblimin rotation of factors. General linear modeling (GLM) in SPSS Complex was used for analyses with metric outcome variables (sum scores as well as single-item variables) and categorical or metric predictors.

In Finland, students from the Swedish-speaking parts of the country were oversampled. These students constitute 17.8% of the Finnish sample of students in the grade used in this study (the grade corresponding to the age group 15 years). An additional challenge was the strong variation in number of study participants across countries, that is, Iceland, with a small population (about 340,000) has the largest sample ( $n = 2244$ ), while Norway, with more than 5.5 million inhabitants, has a much smaller sample ( $n = 876$ ). The data were therefore weighted in such a way that the number of participants for each country is set equal, the proportion of students from the Swedish-speaking part of Finland is reduced, and the number of students in the total sample remains unchanged. When applying these weights to the data, the number of participants from each of the four Nordic countries included in this study equals 1482, and the total number is 5928. The largest weight is for Norway (1.6918) and the smallest weight is for Swedish-speaking students in Finland (0.4352).

In addition to weighting, adjustments for the cluster effect were necessary. On this point we ran into a challenge. The school authorities of Reykjavik, the capital of Iceland, did not permit registration of schools and classes. Reykjavik therefore had to constitute one large cluster. Adjustments for the cluster effect were therefore only partly possible in the data from Iceland. This limitation of the study may have had consequences at a more detailed level of analysis and some of the comparisons across countries. For analyses based on all data combined, it is unlikely that conclusions would change much

even in the absence of this clustering problem. Intraclass correlations for the SWEMWBS and SUS scales with school class as the clustering unit for Iceland did not deviate much from the averages for the Nordic countries combined.

The proportion of students with missing observations on at least one of the items for each scale was 6.3% on the Warwick-Edinburgh Mental Wellbeing scale and 12.4% on the Sense of Unity Scale (weighted numbers). In some analyses students with missing values were included as a separate category. In other analyses, students with missing were excluded.

### 3 | RESULTS

#### 3.1 | Descriptives

As shown in Table 1, the distribution across gender is approximately equal in all countries combined with 49.1% boys and 50.9% girls (weighted numbers). There was no significant difference in gender distributions across countries ( $\chi^2 = 0.461$ ;  $F_{\text{adjusted}} = 0.130$ ;  $df: 2$  and  $711$ ;  $p = .928$ ).

Percentage distributions for all items of the SUS are shown in Table 2. Most items are moderately left-skewed with skewness values ranging from  $-0.59$  to  $-1.18$  (weighted). The number of missing answers varied from 439 (7.4%) to 498 (8.4%) (weighted).

The proportion with responses to all items on the SUS scale was 87.6%, and 12.4% were with missing on at least one item. The number of study participants who had missing on all items of the SUS was 371 (6.3%) (all numbers weighted).

Percentage distributions for all items of the SWEMWBS scale are shown in Table 3. Most items are moderately left-skewed with skewness values ranging from  $-0.39$  to  $-0.77$  (weighted). The number of missing answers varied from 177 (3.0%) to 221 (3.7%) (weighted).

The proportion with valid answers on all items on the SWEMWBS scale was 93.7%, while the proportion with missing on at least one item on the SWEMWBS scale was 6.3%. The number of study participants who had missing on all items was 147 (2.5%).

The mean of the SUS mean score was 3.78 (range 1–5) and the skewness was  $-0.88$ . The SWEMWBS mean score (range 1–5) had a mean of 3.63 and the skewness was  $-0.58$ .

#### 3.2 | Straightlining

The overall proportion of straightliners on the SUS was 22.8% varying from 5.8% among Swedish girls to 46.4% among Finnish boys (Table 4). The variation across countries was significant for both genders (boys:  $\chi^2_{\text{PEARSON}} = 153.354$ ;  $F_{\text{ADJUSTED}} = 37.097$ ;  $df: 2.798$  and  $836.604$ ;  $p < .001$ ). Girls:  $\chi^2_{\text{PEARSON}} = 92.492$ ;  $F_{\text{ADJUSTED}} = 18.945$ ;  $df: 2.907$  and  $892.551$ ;  $p < .001$ ). The proportions of straightliners on the SUS were higher among boys than among girls in all the four countries (Finland:  $\chi^2_{\text{PEARSON}} = 50.849$ ;  $F_{\text{ADJUSTED}} = 39.386$ ;  $df: 1$  and  $73$ ;  $p < .001$ . Iceland:  $\chi^2_{\text{PEARSON}} = 32.310$ ;  $F_{\text{ADJUSTED}} = 25.412$ ;  $df: 1$  and  $100$ ;  $p < .001$ . Norway:  $\chi^2_{\text{PEARSON}} = 15.733$ ;  $F_{\text{ADJUSTED}} = 15.694$ ;  $df: 1$  and  $64$ ;  $p < .001$ . Sweden:  $\chi^2_{\text{PEARSON}} = 32.124$ ;  $F_{\text{ADJUSTED}} = 19.412$ ;  $df: 1$  and  $77$ ;  $p < .001$ ).

The overall proportion of straightliners on the SWEMWBS scale was 18.4%, varying from 5.2% among Norwegian girls to 46.0% among (again) Finnish boys (Table 5). The variation across countries was significant for both genders

**TABLE 1** Respondents by gender and country. Uweighted<sup>a</sup>.

Country	Boys %	Girls %	Total %	$N_{\text{unweighted}}$	$N_{\text{weighted}}^b$
Finland	49.5	50.5	100.0	1146	1473
Iceland	49.4	50.6	100.0	2195	1450
Norway	49.0	51.0	100.0	874	1479
Sweden	48.4	51.6	100.0	1606	1438
Total	49.1	50.9	100.0	5821	5839

Note:  $\chi^2_{\text{Pearson}} = 0.461$ ;  $F_{\text{adjusted}} = 0.130$ ;  $df: 2$  and  $711$ ;  $p = .928$ .

<sup>a</sup>Weighting changes the number of observations for each country, totals across countries, percentages for Finland, and significance testing. Weighting has no effect on percentages for Iceland, Norway, and Sweden.

<sup>b</sup>For all countries the total weighted number of participants is 1482. Due to some missing on gender, the numbers in this table are slightly lower and with some variation across countries.

**TABLE 2** Sense of Unity indicators—descriptives and percentage distributions, weighted.

	Mean	SD	Skewness	Not at all true for me %	Not very true for me %	Neither true nor untrue for me %	Some-what true for me %	Really true for me %	Total %	N <sup>a</sup>	Missing observations N <sup>a</sup>
A strong sense of togetherness	3.67	1.01	-0.64	4.2	5.9	29.8	39.1	21.0	100.0	5430	498
Responsibility for others	3.68	0.98	-0.68	3.7	6.0	28.4	42.4	19.4	100.0	5478	450
It is good to be part of a community	3.96	0.98	-0.95	3.1	4.0	19.9	39.9	33.1	100.0	5473	455
I contribute without expecting anything in return	3.70	0.99	-0.65	3.6	6.1	28.2	40.4	21.7	100.0	5472	456
I mean a lot to others	3.73	1.03	-0.71	3.9	7.4	24.5	40.3	23.9	100.0	5481	447
Others mean a lot to me	4.12	0.97	-1.18	2.7	3.5	15.0	36.8	41.9	100.0	5488	440
I trust most people	3.58	1.10	-0.59	5.4	11.0	24.7	38.1	20.8	100.0	5489	439
We rely upon each other	3.75	1.01	-0.71	3.7	5.8	26.2	40.0	24.3	100.0	5458	470

<sup>a</sup>Numbers of observations weighted.

**TABLE 3** Warwick scale indicators—descriptives and percentage distributions weighted.

	Mean	SD	Skewness	Never %	Rarely %	Every now and then %	Frequently %	All the time %	Total %	n <sup>a</sup>	Missing observations n <sup>a</sup>
I have been feeling optimistic about the future	3.60	1.02	-0.55	3.8	9.8	27.9	39.3	19.1	100.0	5751	177
I have been feeling useful	3.53	0.98	-0.47	3.3	10.3	31.3	40.1	15.0	100.0	5728	200
I have been feeling relaxed	3.44	1.03	-0.39	4.1	13.6	30.8	37.0	14.5	100.0	5738	190
I have been dealing with problems well	3.56	0.98	-0.47	3.1	9.8	30.9	40.1	16.1	100.0	5724	204
I have been thinking clearly	3.68	0.97	-0.57	2.7	8.6	26.9	42.2	19.6	100.0	5707	221
I have been feeling close to other people	3.76	1.01	-0.65	2.8	8.3	23.8	40.2	24.9	100.0	5721	207
I have been able to make up my mind about things	3.84	1.00	-0.77	2.9	6.7	21.9	40.5	27.9	100.0	5712	216

<sup>a</sup>Numbers of observations weighted.

(Boys:  $\chi^2_{PEARSON} = 227.433$ ;  $F_{ADJUSTED} = 50.179$ ;  $df: 2.787$  and  $833.188$ ;  $p < .001$ . Girls:  $\chi^2_{PEARSON} = 93.131$ ;  $F_{ADJUSTED} = 22.962$ ;  $df: 2.822$  and  $866.460$ ;  $p < .001$ ). The proportions of straightliners on the SWEMWBS were higher among boys than among girls in all the four countries (Finland:  $\chi^2_{PEARSON} = 74.339$ ;  $F_{ADJUSTED} = 74.327$ ;  $df: 1$  and  $73$ ;  $p < .001$ . Iceland:  $\chi^2_{PEARSON} = 40.492$ ;  $F_{ADJUSTED} = 48.343$ ;  $df: 1$  and  $100$ ;  $p < .001$ . Norway:  $\chi^2_{PEARSON} = 9.238$ ;  $F_{ADJUSTED} = 8.101$ ;  $df: 1$  and  $64$ ;  $p = .006$ . Sweden:  $\chi^2_{PEARSON} = 52.019$ ;  $F_{ADJUSTED} = 42.848$ ;  $df: 1$  and  $77$ ;  $p < .001$ ).

Among students who straightlined on the SUS, 45.5% straightlined on the SWEMWBS. Among those who did not practice straightlining on the SUS, the proportion practicing straightlining on the SWEMWBS was 10.7%. The difference was significant ( $\chi^2_{PEARSON} = 726.640$ ;  $F_{ADJUSTED} = 486.281$ ;  $df: 1$  and  $317$ ;  $p < .001$ ). The  $\gamma$  coefficient was 0.750 and  $\phi$  was 0.376.

For the SUS, the Cronbach's  $\alpha$  value was .92, highest for Finnish boys (.96) and lowest among Swedish girls (.84). After exclusion of all straightliners, the overall  $\alpha$  for all groups combined was lower (.86), but still high. In all subgroups defined by gender and country, the  $\alpha$  value became smaller after exclusion of straightliners. The effect of excluding all straightliners for the SWEMWBS is the same as shown for the SUS with a reduction in Cronbach's  $\alpha$  for all groups combined is from .90 to .84.

**TABLE 4** Straightlining (identical responses across all eight items) on the Sense of Unity scale by gender and country, weighted and adjusted for cluster effects (school classes). Only students with valid values on all items are included.

Country	Gender	Straightlining			N	N <sub>weighted</sub>	Significance	
		Yes %	No %	Total %			Gender	Country
Finland	Boys	46.4	53.6	100.0	503	642	$\chi^2_{\text{PEARSON}} = 50.849$ $F_{\text{ADJUSTED}} = 39.386$ $df: 1; 73. p < .001$	Boys: $\chi^2_{\text{PEARSON}} = 153.354$ $F_{\text{ADJUSTED}} = 37.097$ $df: 2.798; 836.604$ $p < .001$
	Girls	25.0	75.0	100.0	518	675		
Iceland	Boys	27.3	72.7	100.0	1008	666	$\chi^2_{\text{PEARSON}} = 32.310$ $F_{\text{ADJUSTED}} = 25.412$ $df: 1; 100. p < .001$	
	Girls <sup>a</sup>	16.9	83.1	100.0	1068	705		
Norway	Boys	29.6	70.4	100.0	334	565	$\chi^2_{\text{PEARSON}} = 15.733$ $F_{\text{ADJUSTED}} = 15.694$ $df: 1; 64. p < .001$	Girls: $\chi^2_{\text{PEARSON}} = 92.492$ $F_{\text{ADJUSTED}} = 18.945$ $df: 2.907; 892.551$ $p < .001$
	Girls <sup>a</sup>	16.9	83.1	100.0	355	601		
Sweden	Boys	15.0	85.0	100.0	682	611	$\chi^2_{\text{PEARSON}} = 32.124$ $F_{\text{ADJUSTED}} = 19.412$ $df: 1; 77. p < .001$	
	Girls	5.8	94.2	100.0	737	660		
Total		22.8	77.2	100.0	5205	5124 <sup>b</sup>		

<sup>a</sup>By coincidence, the percentage distributions for Icelandic girls and Norwegian girls are identical.

<sup>b</sup>The sum total of observations in the table is 5124. Due to rounding errors in the weighted numbers, the sum of number of observations across subgroups is 5125.

**TABLE 5** Straightlining (identical responses across all seven items) on the Warwick scale indicators by gender and country. Weighted and adjusted for cluster effects (school classes). Only students with valid values on all items are included.

Country	Gender	Straightlining			N	N <sub>weighted</sub>	Significance	
		Yes %	No %	Total %			Gender	Countries
Finland	Boys	46.0	54.0	100.0	521	663	$\chi^2_{\text{PEARSON}} = 74.339$ $F_{\text{ADJUSTED}} = 74.327$ $df: 1; 73. p < .001$	Boys: $\chi^2_{\text{PEARSON}} = 227.433$ $F_{\text{ADJUSTED}} = 50.179$ $df: 2.787; 833.188$ $p < .001$
	Girls	20.9	79.1	100.0	529	693		
Iceland	Boys	23.2	76.8	100.0	1045	690	$\chi^2_{\text{PEARSON}} = 40.492$ $F_{\text{ADJUSTED}} = 48.343$ $df: 1; 100. p < .001$	
	Girls	12.6	87.4	100.0	1099	726		
Norway	Boys	11.0	89.0	100.0	390	660	$\chi^2_{\text{PEARSON}} = 9.238$ $F_{\text{ADJUSTED}} = 8.101$ $df: 1; 64. p = .006$	Girls: $\chi^2_{\text{PEARSON}} = 93.131$ $F_{\text{ADJUSTED}} = 22.962$ $df: 2.822; 866.460$ $p < .001$
	Girls	5.2	94.8	100.0	421	712		
Sweden	Boys	21.7	78.3	100.0	715	640	$\chi^2_{\text{PEARSON}} = 52.019$ $F_{\text{ADJUSTED}} = 42.848$ $df: 1; 77. p < .001$	
	Girls	8.4	91.6	100.0	782	700		
Total		18.4	81.6	100.0	5502	5485*		

\*The sum total of observations in the table is 5485. Due to rounding errors in the weighted numbers, the sum of number of observations across subgroups is 5484.

### 3.3 | Scale distributions by straightlining

To examine distributions more closely, the mean scores for the two scales (Sense of Unity and Warwick-Edinburgh Mental Wellbeing Scale) were categorized. All values were recoded to the closest of the integers 1 to 5. The categorized variables by straightlining (No/Yes) are shown in Figures 1 and 2.

On both scales, the most extreme scores (1 and 5) are more frequent among straightliners. In both distributions, there are lower percentages around value 4 among straightliners.

On the Sense of Unity score, the means were 3.73 and 3.79 (Wald  $F = 2.381$ ;  $df: 1$  and  $308$ ;  $p = .124$ ), SDs 1.09 and 0.68, and skewness  $-0.75$  and  $-0.81$  for those with and without straightlining. On the SWEMWBS, the corresponding means were 3.81 and 3.59 (Wald  $F = 29.625$ ;  $df: 1$  and  $315$ ;  $p < .001$ ), SDs 1.06 and 0.69, and skewness  $-0.94$  and  $-0.49$  for those with and without straightlining.



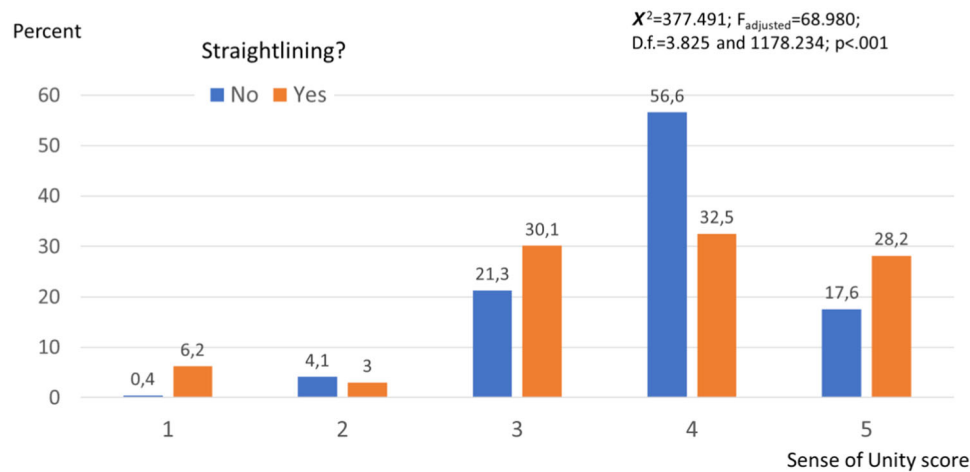


FIGURE 1 Sense of Unity sumscore (categorized) by straightlining on the Sense of Unity scale. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/jad.12066)]

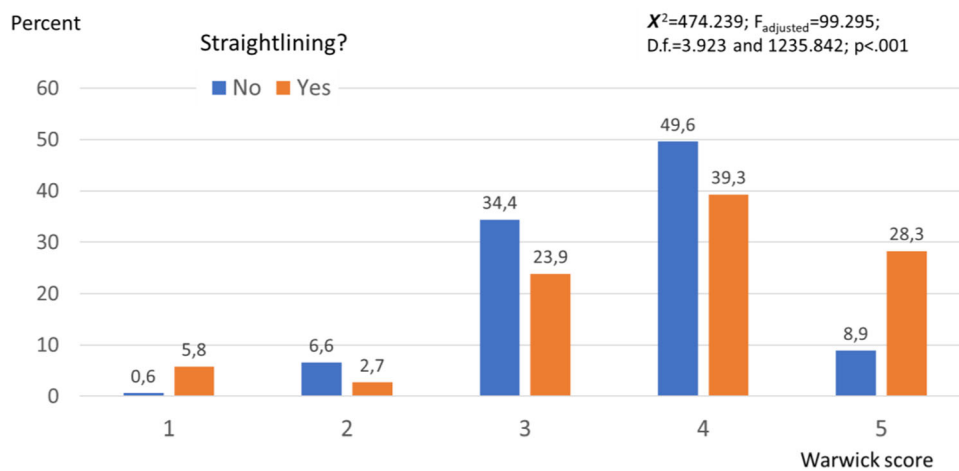


FIGURE 2 Warwick sumscore (categorized) by straightlining on the Warwick scale. [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/jad.12066)]

### 3.4 | Dimensionality

Analyses of dimensionality (factor analysis with principal components extraction and pairwise deletion of missing cases) for the SUS was carried out for each subgroup defined by gender and country, with and without exclusion of straightliners.

For all countries and for both genders, the percentage of variance explained by the first unrotated factor decreased markedly when straightliners were excluded, on average from 62.9% to 51.0%. The largest decrease was for boys in Finland, from 79.4% to 59.6%. In three cases, the number of unrotated factors with eigenvalues higher than 1.00 (Kaiser's criterion) increased from one to two (girls in Iceland, boys in Norway, and boys in Sweden).

Similar analyses were carried out for the SWEMWBS. In all eight subgroups defined by gender and country, the percentage of variance explained by the first unrotated factor decreased, on average from 61.9% to 52.6%. Again, the largest decrease was for boys in Finland where the decrease was from 77.9% to 55.8%. In all analyses of the SWEMWBS only the first factor obtained an eigenvalue higher than 1.00.

### 3.5 | Characteristics of straightliners

In a series of GLM analyses, a selection of outcomes was analyzed against a variable consisting of four groups defined by response sets on the two scales we are dealing with here, the SWEMWBS and the SUS. The four groups were:

(i) no straightlining, (ii) straightlining on one of the scales (SUS or SWEMWBS), (iii) straightlining on both scales, and (iv) missing on at least one item on one of the two scales.

Results for the total sample, weighted as described previously and with adjustments for gender and country, are shown in Table 6. Those with no missing on any SWEMWBS or SUS items and with no straightlining serve as reference. Those who are straightliners on one of the scales had significant deviations on all 13 outcomes, with absolute values ranging from 0.13 to 0.33. Among those who were straightliners on both scales, there were significant deviations on eight of the outcomes with absolute values ranging from 0.13 to 0.37.

All deviations show more favorable scores among straightliners. They do, for instance, score higher on outcomes like Life satisfaction, Self-esteem, Generalized Self-efficacy, Self-rated health, and Peer support, and they score lower on Health complaints and Loneliness.

Those who had missing on at least one item on one of the scales scored slightly (significantly) lower on Sense of Unity and Liking school.

### 3.6 | Associations with external variables

A series of analyses of bivariate associations between ten dependent variables and scores on the SUS as predictor was carried out in four subgroups. Those with: (i) no straightlining on any of the two predictor scales (reference group), (ii) straightlining on one of the scales only, (iii) straightlining on both scales, and (iv) missing on at least one item on one of the scales. Adjustments were made for gender and country and all variables were standardized. An identical set of analyses were carried out with the SWEMWBS as predictor.

An examination of unstandardized regression coefficients revealed that when compared with the reference group, the associations were weaker in 19 out of 20 cases among those who were straightliners on both scales (13 out of 20 with nonoverlapping confidence intervals). In most cases the reduction was rather strong. This was, for instance, the case when Self-esteem was the dependent variable with reductions from 0.52 to 0.14 with the SUS and from 0.68 to 0.17 with the SWEMWBS. Large and significant differences were also observed for Life satisfaction, Self-esteem, Self-efficacy (significant for SWEMWBS scale only), Loneliness, Health complaints, Self-rated health, Likes school (significant for the SWEMWBS only), Relationship with teachers, and Peer support (significant for the SUS only).

**TABLE 6** Selected outcome variables by straightlining adjusted for gender and country analyzed with SPSS complex—general linear modeling. All outcome measures were standardized. Weighting and adjustment for clustering applied.

<b>Straightlining N:</b>	<b>None 3394–3443</b>	<b>One 1007–1034</b>	<b>Both 495–513</b>	<b>At least one item missing 450–815</b>	<b>Significance <i>p</i> &lt;</b>
Sense of Unity	0.00	0.20 <sup>a</sup>	−0.03	−0.14 <sup>b</sup>	0.001
Warwick	0.00	0.33 <sup>a</sup>	0.25 <sup>c</sup>	−0.08	0.001
Life satisfaction	0.00	0.29 <sup>a</sup>	0.33 <sup>a</sup>	0.04	0.001
Self-esteem	0.00	0.31 <sup>a</sup>	0.32 <sup>a</sup>	0.07	0.001
Generalized self-efficacy	0.00	0.18 <sup>a</sup>	0.32 <sup>a</sup>	0.01	0.001
Loneliness <sup>d</sup>	0.00	−0.33 <sup>a</sup>	−0.37 <sup>a</sup>	−0.07	0.001
Complaints	0.00	−0.26 <sup>a</sup>	−0.34 <sup>a</sup>	−0.08	0.001
Health	0.00	0.25 <sup>a</sup>	0.24 <sup>a</sup>	−0.04	0.001
Liking school	0.00	0.13 <sup>a</sup>	0.06	−0.11 <sup>b</sup>	0.001
Relationship with teachers	0.00	0.18 <sup>a</sup>	0.12	−0.05	0.001
Peer support	0.00	0.13 <sup>a</sup>	0.13 <sup>b</sup>	−0.08	0.001
Family support	0.00	0.18 <sup>a</sup>	0.10	−0.01	0.001
Family affluence	0.00	0.13 <sup>a</sup>	0.09	−0.04	0.001

<sup>a</sup>99.9% confidence interval not overlapping 0.00.

<sup>b</sup>95% confidence interval not overlapping 0.00.

<sup>c</sup>99% confidence interval not overlapping 0.00.

<sup>d</sup>Questions on loneliness were not included in the Norwegian questionnaire for 2018. Number of observations across the four groups: 2604; 785; 459; 411.

Among those who were straightliners on one of the scales only, coefficients were weaker than in the reference group in 19 out of 20 cases, and with 11 nonoverlapping confidence intervals.

There is one exception from the rule that straightlining leads to reduced associations between variables. The association between scores on the SWEMWBS and the SUS is significantly higher among those who were straightliners on both scales (coeff. = 0.734) than in the reference group (coeff. = 0.524).

### 3.7 | Associations between external variables

In a series of bivariate analyses with health complaints as dependent variable and eight selected predictors (Self-esteem, Generalized self-efficacy, Loneliness, Liking school, Teacher support, Peer support, Family support, Family affluence) and adjustments for Gender and Country, the mean unstandardized regression coefficient in the reference group was 0.281. Among those with straightlining on one of the two scales (Sense of Unity and Warwick-Edinburgh), the mean coefficient was 0.231. Among those with straightlining on both scales, the mean coefficient was 0.210. Identical analyses with Life Satisfaction as the dependent variable gave a similar pattern of correlations: 0.354, 0.292, and 0.273. Analyses with Self-rated health as the dependent variable also gave a similar pattern of correlations: 0.258, 0.226, and 0.217.

## 4 | DISCUSSION

Straightlining means that the same answer is provided across all items within a multi-item scale. To distinguish between valid and invalid straightliners at the level of individual study participants is difficult or impossible. There is no easy way to distinguish between those who have invested adequate time and reflection when responding, and gave the same answer to all items, and those who have answered quickly and superficially and chose the same response across all items to save time and effort or simply lacked motivation. When the proportion of straightliners becomes high, however, we may assume that the quality of responses may be reduced for at least a proportion of the respondents. Differences in the occurrence of straightlining across subgroups may also serve as an indicator that the quality of answers differs across groups.

Straightlining turned out to be surprisingly prevalent, with considerable variation across subgroups defined by gender and country. Among those who had valid values on all items on both scales, 31.1% of the students in all the four countries combined (and 60.1% of boys in Finland) were straightliners on one or both scales.

In the present data, the overall prevalence of straightliners is 22.8% on the SUS and 18.4% on the SWEMWBS. These numbers are sufficiently high to indicate superficial responses by a substantial proportion of students. There is considerable variation across subgroups defined by gender and country. Among boys in Finland, the proportion of straightliners was as high as 46.4% on the SUS. Among girls in Sweden, the proportion was only 5.8%. Among boys in Finland, the proportion of straightliners on the SWEMWBS was 46.0%. Among girls in Norway the proportion was only 5.2%. It is difficult to find good explanations why the proportions are so high in some of the groups beyond the assumption that some students were quick and superficial when filling in the questionnaire. If all students had been highly motivated, and not bored by the length of the questionnaires, and if all students had examined and responded to the questions after careful consideration of the alternative responses, it is unlikely that we would have straightlining among more than a small fraction of respondents.

Straightlining is far more prevalent among boys than among girls on both scales and in all countries. This is not at all surprising. A consistent finding in research on personality traits among adolescents is higher levels of conscientiousness among girls (De Bolle et al., 2015). This gender difference is therefore consistent with the assumption that superficial cognitive processing (less conscientious responding) contributes to straightlining. Furthermore, studies have shown that girls in general produce slightly more reliable answers and higher quality data than boys do (Omran et al., 2019). Also, international reading assessment shows that girls are more engaged and better readers than boys (Brozo et al., 2014).

On a number of indicators, straightlining students score more favorably than non-straightlining students. Straightlining does not seem to primarily be a strategy chosen by students who lack abilities and resources to cope with demanding tasks. On the contrary, it may be a strategy preferred by students who are able to provide high-quality responses and who feel good about themselves, but also who feel free to take shortcuts and find easy solutions. This is, for instance, indicated by their higher mean levels of Life satisfaction, Self-esteem, and Generalized Self-efficacy. Studies have shown that the tendency to speed when responding to survey questions is related to higher proportions of straightlining (Zhang & Conrad, 2014).

Students who are bored or exhausted are expected to fall into the pattern of providing identical responses to all items more easily than those who are less bored and less exhausted. Previous research has shown that straightlining is more common in scales that are located towards the end of questionnaires (Herzog & Bachman, 1981). In the Finnish questionnaire, the SWEMWBS items come as number 213–219 out of 232 items, and the Sense of Unity items are items number 220–227 (Table 7). Therefore, it does not come as a surprise that Finland has much higher prevalence of straightlining than the other countries in both genders on both scales.

**TABLE 7** Location of the scales (the Warwick-Edinburgh Mental Well-being scale and the Sense of Unity Scale) within the questionnaire by country.

	The Sense of Unity Scale	The Warwick-Edinburgh Mental Well-being Scale	Total number of items in the questionnaire
Finland	Items 220–227	Items 213–219	232
Iceland	Items 36–43	Items 29–35	181
Norway	Items 181–188	Items 22–28	233
Sweden	Items 91–98	Items 102–108	175

The remaining differences across countries are less straightforward to explain. In the Swedish questionnaire, the SUS appeared halfway through the questionnaire (items 91–98 out of 175 items) (see Table 7 for an overview of scale locations within questionnaires). Still the prevalence of response sets among Swedish girls was considerably lower than for the other countries. This could be explained by the fact that the wordy introduction to the SUS questions was left out in the Swedish questionnaire. In the Norwegian questionnaire, the Sense of Unity items appeared at the beginning of the questionnaire (items 22–28 out of 233) and the SWEMWBS items appeared rather late (items 181–188). Still, the prevalence of straightlining was much higher on the SUS. There are obviously factors beyond levels of boredom and exhaustion caused by location in the questionnaires that determine the tendency to respond with straightlining.

Except from Finland, where the questionnaire was offered in two different languages, languages follow countries. This means that differences across countries to a very large extent also are differences between languages. In Finland, the Swedish-speaking students constitute only a small minority (6%) of the students. This group is too small for separate analyses, and will only have marginal impact on estimates based on the Finnish data. The two Norwegian languages are too similar to represent any challenge with regard to translation.

Previous research carried out among young adults shows that straightlining is not associated with the mode of survey administration/survey method (Kim et al., 2018). However, in this study, Sweden was the only country using pen and paper survey, and they also, for both genders, had the lowest percentages of straightlining on the SUS and the second lowest on the SWEMWBS.

Omrani and associates (2019) have recommended that the number of words of questions in surveys among adolescents aged 11–16 should not exceed 20. The number of words used in the introduction to the SUS is 65 (English language version of the scale). In Sweden, the rather wordy introduction to the SUS was dropped. As mentioned above, Swedish students of both genders had much lower percentages of straightlining on this scale than was the case with students from the other Nordic countries.

The tendency to straightline might be higher during the end of school hours and at the end of school weeks. Since the data collections could take place any day in the week and at any time during school hours, in this study there is no way to find out to what extent these factors might have influenced the tendency to straightline.

If we restrict our statistical analyses to the straightliners only, correlations between variables within each of the scales are all 1.00. Cronbach's  $\alpha$  values will also be 1.00. In data sets with high proportion of straightliners, the  $\alpha$  coefficients will therefore increase. Since  $\alpha$  is a measure of the internal consistency of a scale, and since high internal consistency is regarded to be a quality criterion, high  $\alpha$  values are usually taken as an indicator of high scale quality. When a high  $\alpha$  value stems from straightlining, this is a questionable conclusion. If we exclude all straightliners from our data, the overall  $\alpha$  values for the SUS decrease from .92 to .86. On the SWEMWBS,  $\alpha$  decreases from .90 to .84. In the group with the highest proportion of straightliners (boys, Finland), the  $\alpha$  values decrease from .96 to .90 for the SUS and from .95 to .83 for the SWEMWBS. All these  $\alpha$  values are sufficiently high to be regarded as most satisfactory, even after the exclusion of the straightlining students. Provided that some proportion of straightlining is "invalid," it is obvious that straightlining contributes to inflating the  $\alpha$  values.

Since straightlining leads to higher correlations among items within a scale, it does not come as a surprise that this has an effect of analyses of dimensionality of scales, such as principal components and factor analysis. A high proportion of invalid straightlining will tend to increase the eigenvalue of the first unrotated factor or component and may potentially lead to incorrect conclusions regarding number of components or factors.

Highly skewed indicators represent a challenge when analyzing associations between variables and analyses which are based on correlations like principal components- and factor analysis. Relevant remedies are score transformations, use of polychoric correlations, and robust estimators. All indicators in the SUS and SWEMWBS are left-skewed, and so are therefore the mean scores of both scales. On the SUS, skewness is actually slightly less pronounced among the SUS straightliners. On the SWEMWBS, however, skewness is much stronger among the SWEMWBS straightliners. This indicates that in some cases a high proportion of straightliners in a set of data may amplify the statistical problems associated with skewness.

Among straightlining students, bivariate associations between Sense of Unity scores and scores on the SWEMWBS and a set of variables which in this context were treated as outcomes tended to be weaker than in what we defined as the reference group (non-straightliners). This indicates more arbitrary responses among straightlining students. However, all associations were still in the same direction, and in most cases the differences were not dramatic. Still our findings indicate that a substantial proportion of the straightlining students were “invalid” straightliners. They may not have invested sufficient efforts into reading the questions carefully and providing high-quality responses. However, occurrence of straightlining should not automatically be interpreted as an indication of poor response quality in general. Closer inspections of differences in patterns of associations between straightliners and non-straightliners are recommended.

The strong association between the scores on our two scales among those who were straightliners on both scales deviates from the general finding that straightlining is associated with lower associations. When interpreting this finding, it is important to keep in mind that this association is identical to the associations between the two first items in the two scales. There may be a tendency among straightliners that those who chose a high value as their first response on one scale also chose a high value as a first response on the second scale, while other students may have tended to start out with low values on both scales.

If straightlining was a more general indicator of poor-quality responses to the questionnaire, we would expect associations among any set of variables to differ between subgroups defined by straightlining. If more arbitrary responses were generally more common among straightliners, associations between external variables would be expected to be weaker, in other words “washed out” by randomness. In three series of bivariate analyses with adjustment for Gender and Country, associations differed as expected, but the differences were not dramatic. In other words, straightlining is probably not a strong indicator of poor data quality beyond the scales where straightlining takes place.

#### 4.1 | Strengths and limitations

The most important strengths of this study are the use of the same scales across four different countries and that the total sample size was appropriate. When the purpose is to examine aspects of streamlining, it was necessary to have scales with several items, multiple response categories, identical response categories, and the same order of categories within each scale. Both scales used in the present analyses measure rather abstract concepts and may be cognitively demanding for study participants. To have two different scales of this kind within one single set of data allows examination of streamlining on only one scale and on both scales simultaneously. All these properties of the present data set make it streamlined for the purpose of studying straightlining.

The main limitations of this study are low response rates in all countries and a rather small sample in Norway. The location of the scales within the questionnaire varied a lot across countries. And in one country (Sweden), most of the text introducing the SUS had been removed. On the SUS, there were also some minor differences in the wording of response categories.

## 5 | CONCLUSION

Straightlining has a number of potential consequences which need to be considered when analyzing data and drawing conclusions from studies. Correlations within scales and Cronbach's  $\alpha$  values tend to be inflated. Researchers should be aware that to the extent that high correlations and  $\alpha$  values are caused by straightlining; high values do reflect not only high quality of the data but also limitations of data quality. When analyzing dimensionality, large groups of respondents straightlining tend to favor unidimensionality. Associations between mean scores based on scales where straightlining occurs and other variables may be lower than would be the case with no straightlining among study participants, and even the strength associations among external variables may be influenced (notably, reduced). When researchers use data with scales where there is a high risk of straightlining (scales with many items, many response categories, all response categories in the same order, cognitively demanding scales, scales that come towards the end of lengthy questionnaires), the prevalence of straightlining should be reported and data analyzed to examine consequences. When data collections are prepared, strategies to motivate participants to spend sufficient time and efforts are important. Researchers working with surveys among adolescents should be aware that straightlining is not necessarily an indication of lack of cognitive and other resources. Quite the contrary. Straightlining tends to be an easy solution and shortcut among those with high levels of self-efficacy, self-esteem, and satisfaction with life.

Cognitively demanding scales which come towards the end of a long questionnaire increase the prevalence of straightlining students. Researchers should carefully examine the acceptable length of questionnaires and order of questions and scales, keeping in mind that the average reading performance in some countries has decreased (Schleicher, 2018).



## ACKNOWLEDGMENTS

Health Behaviour in School-aged Children (HBSC) is an international study carried out in collaboration with WHO/EURO. The International Coordinator for the 2017/2018 survey was Jo Inchley (University of Glasgow). The Data Bank Manager was Professor Oddrun Samdal (University of Bergen). The surveys from this study were conducted by the following principal investigators in the four countries: Finland, Jorma Tynjälä; Iceland, Ársæll Már Arnarsson; Norway, Oddrun Samdal; Sweden, Petra Löfstedt. This research was conducted within the research project “Positive mental health among adolescents in the Nordic countries,” funded by the Public Health Agency of Sweden, grant number 0493-2020-2.3.2. Anne-Siri Fismen is a postdoctoral fellow in the CO-CREATE project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 774210.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data file used for this study cannot be deposited online. The HBSC Data Management Centre organizes and distributes international HBSC data. The data file (HBSC 2018) is open access and readers can apply for access and permission to analyze data. Application form and general information are available at [www.hbsc.org](http://www.hbsc.org).

## ETHICS STATEMENT

The study was conducted according to the Guidelines of the Declaration of Helsinki. In Finland, ethical clearance of the study was approved by the Finnish Teachers' Union and the Finnish National Board of Education. In addition, approval for the HBSC-study participation was requested through letters to the school principals. In Iceland, the study was approved by the National Data Protection Agency and the University of Iceland Ethics Committee. Furthermore, permission was sought from education authorities in each municipality as well as principals in every participating school. In Norway, the Privacy Ombudsman at the Norwegian Social Science Data Services confirmed that the study complied with privacy and confidentiality requirements. The Norwegian Western Regional Ethical Committee decided that ethical clearance was not needed. The Swedish study using HBSC data is deemed exempt from human subject research review by the Regional Ethical Review Board in Stockholm. Oral and written information on the confidentiality of study participants' responses were provided and participation was confidential and voluntary. Informed consent of parents (guardians) and adolescents participating in the study was required in Norway. Finland, Iceland, and Sweden used passive consent.

## ORCID

Leif Edvard Aarø  <http://orcid.org/0000-0001-5027-3611>

## REFERENCES

- Brozo, W., Sulkunen, S., Shiel, G., Garbe, C., Pandian, A., & Valtin, R. (2014). Reading, gender, and engagement: lessons from five PISA countries. *Journal of Adolescent & Adult Literacy*, 57(1), 584–593. <https://doi.org/10.1002/jaal.291>
- Cantril, H. (1965). *The pattern of human concern*. Rutgers University Press.
- Canty-Mitchell, J., & Zimet, G. D. (2000). Psychometric properties of multidimensional scale of perceived social support in urban adolescents. *American Journal of Community Psychology*, 28, 391–400. <https://doi.org/10.1023/A:1005109522457>
- Currie, C. E., Elton, R. A., Todd, J., & Platt, S. (1997). Indicators of socioeconomic status for adolescents: The WHO Health Behaviour in School-Aged Children Survey. *Health Education Research*, 12(3), 385–397.
- De Bolle, M., De Fruyt, F., McCrae, R. R., Löckenhoff, C. E., Costa, P. T., Aguilar-Vafaie, M. E., Ahn, C. K., Ahn, H. N., Alcalay, L., Allik, J., Avdeyeva, T. V., Bratko, D., Brunner-Sciara, M., Cain, T. R., Chan, W., Chittcharat, N., Crawford, J. T., Fehr, R., Ficková, E., ... Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. *Journal of Personality and Social Psychology*, 108(1), 171–185.
- Edwards, L. (2010). Measuring perceived social support in Mexican American youth: Psychometric properties of the multidimensional scale of perceived social support. *Hispanic Journal of Behavioral Sciences*, 26(2), 187–194.
- Eriksson, C., Arnarsson, Á. M., Damsgaard, M. T., Löfstedt, P., Potrebny, T., Suominen, S., Thorsteinsson, E. B., Torsheim, T., Välimaa, R., & Due, P. (2019). Towards enhancing research on adolescent positive mental health. *Nordic Welfare Research*, 4(2), 113–128.
- Hartley, J. E., Levin, K., & Currie, C. (2016). A new version of the HBSC Family Affluence Scale—FAS III: Scottish Qualitative Findings from the International FAS Development Study. *Child Indicators Research*, 9, 233–245.
- Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45(4), 549–559.
- Hetland, J., Torsheim, T., & Aarø, L. E. (2002). Subjective health complaints in adolescence: Dimensional structure and variation across gender and age. *Scandinavian Journal of Public Health*, 30(3), 223–230.
- Inchley, J. C. D., Budisavljevic, S., Torsheim, T., Jästad, A., & Cosma, A. (2020). *Spotlight on adolescent health and well-being*. Findings from the 2017/2018 Health Behaviour in School-aged Children (HBSC) survey in Europe and Canada. International report. Volume 3: Methodology annex. WHO Regional Office for Europe.
- Inchley, J. C. D., Cosma, A., & Samdal, O. (Eds.). (2018). *Health Behaviour in School-aged Children (HBSC) Study Protocol: Background, methodology and mandatory items for the 2017/18 survey*. University of St. Andrews, Child and Adolescent Health Research Unit (CAHRU).

- Jovanović, V. (2016). The validity of the satisfaction with life scale in adolescents and a comparison with single-item life satisfaction measures: A preliminary study. *Quality of Life Research*, 25, 3173–3180. <https://doi.org/10.1007/s11136-016-1331-5>
- Kim, Y., Dykema, J., Stephenson, J., Black, P., & Moberg, D. P. (2018). Straightlining: overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review*, 37(2), 214–233. <https://doi.org/10.1177/0894439317752406>
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.
- Levin, K. A., & Currie, C. (2014). Reliability and validity of an adapted version of the Cantril Ladder for use with adolescent samples. *Social Indicators Research*, 119, 1047–1063.
- Loosveldt, G., & Beullens, K. (2017). Interviewer effects on non-differentiation and straightlining in the European Social Survey. *Journal of Official Statistics*, 33(2), 409–426.
- Lyyra, N., Thorsteinsson, E. B., Eriksson, C., Löfstedt, P., Madsen, K., Tolvanen, A., & Välimaa, R. (2021). The association between loneliness, mental well-being, and self-esteem among adolescents in four Nordic countries. *International Journal of Environmental Research and Public Health*, 18(14), 7405. <https://doi.org/10.3390/ijerph18147405>
- Meilstrup, C., Holstein, B. E., Nielsen, L., Due, P., & Koushede, V. (2020). Self-efficacy and social competence reduce socioeconomic inequality in emotional symptoms: A social epidemiological study of schoolchildren. *European Journal of Public Health*, 30(1), 80–85. <https://doi.org/10.1093/eurpub/ckz058>
- Moser, C. A., & Kalton, G. (1971). *Survey methods in social investigation*. Heinemann.
- Nielsen, L., Stewart-Brown, S., Vinther-Larsen, M., Meilstrup, C., Holstein, B. E., & Koushede, V. (2016). High and low levels of positive mental health: are there socioeconomic differences among adolescents? *Journal of Public Mental Health*, 15, 37–49.
- Omrani, A., Wakefield-Scurr, J., Smith, J., & Brown, N. (2019). Survey development for adolescents aged 11–16 years: A developmental science based guide. *Adolescent Research Review*, 4, 329–340.
- Reuning, K., & Plutzer, E. (2020). Valid vs. invalid straightlining: The complex relationship between straightlining and data quality. *Survey Research Methods*, 14(5), 439–459.
- Ringdal, R., Eilertsen, M.-E. B., Bjørnsen, H. N., Espenes, G. A., & Moksnes, U. K. (2017). Validation of two versions of the Warwick-Edinburgh Mental Well-Being Scale among Norwegian adolescents. *Scandinavian Journal of Public Health*, 46(7), 718–725.
- Samdal, O., Moreno, C., Morgan, A., Matos, M. G., Banan, A., & Wold, B. (2016). *Stimulating Adolescent Life skills through Unity and Drive (SALUD)*. Poster at HBSC meeting, Luxembourg, Nov. 3, 2016.
- Schleicher, A. (2018). *PISA 2018. Insights and interpretations*. OECD. <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- Schwarzer, R. (1992). Self-efficacy in the adoption and maintenance of health behaviors: Theoretical approaches and a new model. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action* (pp. 217–243). Hemisphere.
- Steinberg, L., Graham, S., O'Brian, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age differences in future orientation and discounting. *Child Development*, 80(1), 28–44.
- Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes*, 7, 15. <https://doi.org/10.1186/1477-7525-7-15>
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., & Scott Weich, J. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, 2007, 63. <https://doi.org/10.1186/1477-7525-5-63>
- Torsheim, T., Samdal, O., Rasmussen, M., Freeman, J., Griebler, R., & Dur, W. (2012). Cross-national measurement invariance of the Teacher and Classmate Support Scale. *Social Indicator Research*, 105(1), 145–160.
- Torsheim, T., Wold, B., & Samdal, O. (2000). The Teacher and Classmate Support scale—factor structure, test-retest reliability and validity in samples of 13- and 15-year-old adolescents. *School Psychology International*, 21(2), 195–212.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195–217.
- Villar, A. (2008). Response bias. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (Vol. 2, pp. 751–753). Sage.
- Wahlström, J., Modin, B., Svensson, J., Löfstedt, P., & Brolin Läftman, S. (2021). Sense of unity and self-reported health among 15-year-olds: Findings from the Swedish 2017/18 Health Behavior in school-aged Children Study. *International Journal of Public Health*, 66(36), 621964. <https://doi.org/10.3389/ijph.2021.621964>
- Yan, T. (2008). Nondifferentiation. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (Vol. 2, pp. 520–521). Sage.
- Zhang, C., & Conrad, F. (2014). Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127–135. <https://doi.org/10.18148/srm/2014.v8i2.5453>
- Zimet, G., & Grodaon, K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment*, 52(1), 30–41.
- Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment*, 52(1), 30–41.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Aarø, L. E., Fismen, A.-S., Wold, B., Skogen, J. C., Torsheim, T., Arnarsson, Á. M., Lyyra, N., Löfstedt, P., & Eriksson, C. (2022). Nordic adolescents responding to demanding survey scales in boring contexts: Examining straightlining. *Journal of Adolescence*, 94, 829–843. <https://doi.org/10.1002/jad.12066>